DOCUMENT RESUME

ED 032 610 24 CG 004 713

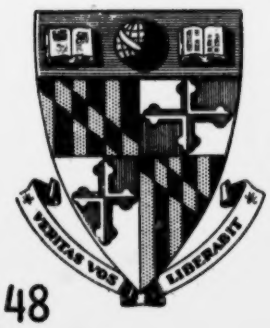
By-Blum, Zahava D.; And Others

A Method for the Collection and Analysis of Retrospective Life Histories. Report No. 48.

Johns Hopkins Univ., Baltimore, Md. Center for the Study of Social Organization of Schools.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-6-1610
Pub Date Jul 69
Grant-OEG-2-7-06160-0207
Note-37p.
EDRS Price MF-$0.25 HC-$1.95
Descriptors-Computers, Educational Research, *Longitudinal Studies, Occupational Mobility, Racial Differences, *Social Class, *Socioeconomic Background Surveys

To achieve the goal of "equality of opportunity" and socioeconomic well being in American society there is the need to know what type of directed social change is appropriate. The purpose of this research effort was to examine empirical data on how social groups and individual households achieve social mobility in order to identify alternative intervention points. Social mobility is seen primarily in socioeconomic terms with a strong but not exclusive, emphasis on the role played by educational attainments in the mobility of individual households and social groups. A current emphasis of our research is an investigation of the occurrence (or lack of occurrence) of certain events as the individual passes through his own life cycel, i.e., a study of intragenerational mobility. Using survey research methods, retrospective life histories have been obtained for a national sample of the noninstitutionalized population of males 30-39 years of age residing in households in the United States and a similar sample of Negro males. This paper, while not intended to be a comprehensive manual, discusses the general approach developed in the project, illustrates many of the procedures, and gives examples from the survey data. (Author/KJ)

ED0 32610



REPORT NO. 48

THE JOHNS HOPKINS UNIVERSITY

THE CENTER FOR THE STUDY OF SOCIAL ORGANIZATION OF SCHOOLS

A METHOD FOR THE COLLECTION AND ANALYSIS OF RETROSPECTIVE LIFE HISTORIES

ZAHAVA D. BLUM

NANCY L. KARWEIT

AAGE B. SØRENSEN

JULY, 1969

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

BR. 6-1610
Report # 48
PA-24
OE/BR

A METHOD FOR THE COLLECTION AND ANALYSIS OF RETROSPECTIVE LIFE HISTORIES

OEG-2-7-06160-0207

Zahava D. Blum
Nancy L. Karweit
Aage B. Sørensen

The Johns Hopkins University

Published by the Center for the Study of Social Organization of Schools and the Learning Process, The Johns Hopkins University, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

Prepared for presentation to the Russell Sage Foundation Seminar on Models and Measurement of Social Change, May, 1969

The Johns Hopkins University
Baltimore, Maryland

ABSTRACT

To achieve the goal of "equality of opportunity" and socio-economic well being in American society there is the need to know what type of directed social change is appropriate. The purpose of our research effort is to examine empirical data on how social groups and individual households achieve social mobility in order to identify alternative intervention points. Social mobility is seen primarily in socio-economic terms (occupation and income) with a strong, but not exclusive, emphasis on the role played by educational attainments in the mobility of individual households and social groups.

A current emphasis of our research is an investigation of the occurence (or lack of occurrence) of certain events as the individual passes through his own life cycle, i.e., a study of intragenerational mobility. Using survey research methods, retrospective life histories have been obtained for a national sample of the noninstitutionalized population of males 30-39 years of age residing in households in the United States and a similar sample of Negro males.

The feasibility of collecting retrospective life history data from national samples and of efficiently coding, storing and gaining computer access to the resultant information has been demostrated by this research effort. This paper, while not intended to be a comprehensive manual, discusses the general approach developed in the project, illustrates many of the procedures, and gives examples from the survey data.

CONTENTS

Introduction	1
Procedure	5
Questionnaire	5
Field Work	8
Evaluation	15
Processing the Data	17
Machine Processing	22
Summary	26

Figures

1	7-8
2	20
3	21
4	24

Tables

1	9-12
2	27
3	28
4	29
5	30

Bibliography	31
Appendix A	A-1

The broad, long-range commitment of our research group at The Johns Hopkins University is to the development of a system of social accounts and to considerations of models and data systems necessary for monitoring social change.[2] When we speak of the development of a system of social accounts we are conscious of the implications that such a system can have for purposive social change.[3] Any discussion of such social change must imply some ideas at the outset of the desired or optimum state toward which change might be directed. The vision of the future to which we are ideologically committed might be termed "equality of opportunity" for all members of our society.

In spite of disagreement about the exact nature of an ideal future, we believe there <u>is</u> agreement on the basic phenomena which should be studied in order to understand and eventually monitor the individual's position in society. However, we are confronted with the realization that knowledge of the processes which "make a difference" in the lives of individuals and sub-groups in our society is sufficiently meager to warrant a careful, incremental approach to our long-term goal.

In our deliberations, Rossi suggested the following alternative (and narrower) foci as being appropriate entry points into actual research:

---

[2]The authors of this paper are, respectively, project co-director, associate research scientist, and research assistant of the Social Accounts Research Group. This project is under the direction of Peter H. Rossi, James S. Coleman, and Zahava D. Blum. We would like to acknowledge the enthusiastic assistance of the following graduate students in all phases of the project: Richard Conviser, David Grafstein, Daniel Hadary, John Kervin, and Wen-hsiung Kuo.

[3]A preliminary theoretical framework within which a system can be developed is presented in Coleman (1969).

Investigation of:

1) processes of intergenerational mobility,

2) processes of intragenerational mobility,

3) "community" influences on social mobility and change, or

4) computer simulation of mobility processes.

These are alternatives only for purposes of allocating current efforts and resources. The recent work by Blau and Duncan (1967) has provided a firm foundation for understanding intergenerational mobility, while we have limited faith, at this stage, in our ability to conceptualize (and operationalize) the needed research in the study of "community" influences. Finally, since computer simulation is an unrealistic investment of energy without a better grasp of the various processes mentioned above, we have decided to focus at present on intragenerational mobility, believing that we could most effectively make a contribution in that area.

This decision and our belief in "equality of opportunity" led us to emphasize racial differences in intragenerational processes. Consequently, we designed a study which would allow comparisons between blacks and whites in the United States. Preliminary discussions suggested that our analytic interests could best be served by interview data from approximately equal numbers of blacks and whites.[4] In addition, we felt that life history data in the form of longitudinal or retrospective material was the only appropriate

---

[4]Given our samples (See Appendix A), we can perform our analysis using either of two comparisons: black and nonblack or blacks and a national sample of the population.

empirical base with which to approach some of the substantive interests sketched above.

The need for longitudinal data is often noted in the sociological literature of the past few decades. The common approach, due to serious technical difficulties, has been the resort to cross-sectional data in most empirical analysis; attempts to deal with longitudinal materials have been mostly social-anthropological in nature. We are not implying that attempts to overcome the technical and methodological difficulties inherent in longitudinal data have not been made. Certainly, cohort analysis, panel studies, time-series analysis, etc., are all notable efforts in that direction. The fact does remain, however, that a satisfactory methodology for the analysis of longitudinal data does not now exist in sociology. It is our belief that recent developments in systems of large-scale data collection and advances in computer technology make a confrontation of the "longitudinal problem" less formidable than in the past.

It is the aim of this paper to present a method for the collection and processing of life history material developed in our research project, as well as give some illustrations of the type of analytic flexibility the approach allows.[5] It is our impression that the set of procedures we have developed will enable us, and other investigators, to take full advantage of the analytic possibilities inherent in similar data.

---

[5]This paper is not intended to be a comprehensive manual; further inquiries about procedures and computer programs are welcomed by the authors.

A recent paper in Behavioral Science (Balan et al, 1969) was directed toward encouraging social scientists to collect and analyze life history material. It would be unjust on our part to imply that communication with the authors of the above cited work was not helpful to us. However, we feel that our proposals will help simplify the technical problems encountered in dealing with life history data, and that these proposals have broader analytic implications. For purposes of comparison, the remainder of this paper follows the Balan et al presentation, henceforth called the Monterrey study.

Whereas the Monterrey study defined the universe as the resident male population, between 21 and 60 years of age of the Metropolitan Area of Monterrey, the target population for our study is the total population of males 30-39 years of age in 1968 residing in households in the United States.[6] Budgetary considerations limited our sample size and we decided to restrict ourselves to one age group. We chose men 30-39 for several reasons. First, the majority of our respondents entered the labor force following World War II; second, they will have completed their education and had some labor force experience. Before proceeding to the technical aspects of the study, it should be noted that our design makes possible the use of two somewhat disparate analytic approaches. We intend to utilize the methodology in the sociological tradition of social mobility research. Also, we have been

---

[6]To the best of our knowledge, our study is the first to collect detailed retrospective histories for a national sample in this country.

intrigued with the possibility of taking the approach from economics most generally described as studies of "human capital" and applying some of the conceptualizations and techniques to individual data.

PROCEDURE

Questionnaire

Experimentation with a number of formats led to the development of an instrument which systematically records information from the respondent's age 14 to the present on thirteen major variables.[7] The life history section of the questionnaire assigns columns to each of the variables and rows to each year covered by the survey (1943-1968). In addition the questionnaire contains a number of cross-sectional or static variables (e.g., parental education and occupation at respondent's age 14, ethnic background, religion, etc.).

Whereas the Monterrey study defined the smallest period of time in eliciting information for various aspects of the respondent's life as <u>six</u> months or more, we have defined the <u>month</u> as our smallest unit. As illustrated in Figure 1, the actual calendar month in which a specific change took place is recorded. As will be discussed later, the month is preserved as the analytic unit in the Hopkins study, while the Monterrey study recorded activities in <u>years</u>. The latter approach forces all changes to take place

---

[7] A summary of the information is presented in Table 1. Two sections of a completed questionnaire are presented on the following pages (Figure 1).

INTERVIEWER: ALWAYS INDICATE STOPPING POINT BY END OF ARROW.

Year	Age	1. FULL TIME EDUCATION (a) Month	(b) Name and/or type of school	(c) Degree/highest gr.comp.	2. FULL TIME EMPLOYMENT OR UNEMPLOYMENT (Ask 11-a "Support," for unempl.) (a) Month	(b) Occupation	(c) Industry	Wages (d) In $ Beg. Amt.	(d) In $ End'g Amt.	(e) In kind F=food H=hous'g	(f) Hrs/wk	(g) Left Job: O=own decision N=not own decision	(h) At termination: H=had new job K=knew of job N=neither	(i) Got Job thru: Fr=Friends Fa=Family Pu=Pub.ag. Pr=Priv.ag. A=Ads O=Other	(j) On-the-Job Training (incl. apprentice trng.) How long? Never? ☐
1943	14		Barker Elem.												
1944	15	June Sept	Barker Cons. HS	8 9											
1945	16	June Sept	St. Louis #5 Public HS	10											
1946	17	Sept	↓ Oct	10	Nov Dec	Unemployed (see note p. 3) Attendant	Gas Station	.75 hr			40			A	
1947	18					↓	↓		.75hr			N	N		
1948	19				Apr-July	Stock Clerk Army (see 5.)	Dept. Store	.80 hr	.80		40	N	N	Pu	
1949	20														
1950	21														
1951	22				Feb	Driver	City Bus Co.	1.10 hr			40			A	Driver Training 3 weeks
1952	23					↓	↓								
1953	24	Sept	City College of New York		Aug				1.30 hr		40	O	N (see note)		
1954	25		↓												
1955	26	June	↓	BA (business)	June	Salesman	Levin Bros. Soap Mfg.	$6,000 yr			42			O	Management Training Program - 6 weeks
1956	27						↓								
1957	28														
1958	29				Aug	Area Supervisor	O. Sedar Brush Co.	$8,250 yr	$7,500 yr		40 40-45	O	H	O	
1959	30						↓								
1960	31														
1961	32														
1962	33														
1963	34														
1964	35				Feb	↓			$11,000 yr		40-45	N leave of absence for illness	N		
1965	36				March	Office Manager		$12,000 yr			45-50			back to same job	
1966	37					↓									
1967	38														
1968	39														
Current						Office Manager	O. Sedar Brush Co.	$13,000 yr			45-50				

Figure 1

INTERVIEWER: ALWAYS INDICATE STOPPING POINT BY END OF ARROW.

Year	7. FAMILY HISTORY (a) Month & Marital Status	(b) Wife's age at marriage	(c) Children M=male F=female Mo. of birth	(d) Birth control dur. mar. Never? ☐	8. WIFE'S EDUCATION Degree or highest gr. comp. at mar. & since	9. WIFE'S EMPLOYMENT (during marriage to R.): Never? ☐ (a) Month	(b) Occupation	(c) Industry	(d) Wages	(e) Hrs/wk	10. HOUSEHOLD COMPOSITION: Relationships	11. INCOME ADEQUACY: S=Save C=Comfort, no save M=Manage D=Need outside help	(a) Public Support Share of total income A=All M=More than ½ H=About half L=Less than ½ N=None	Kind of Support
1943											Mother, Step-father, 2 Bros., Aunt	M		
1944														
1945											(June) Mother, Step-father, 2 Brothers			
1946	Nov married	16		no	10th gr						(Nov) Wife and above family	D (family)		
1947			M-July								(June) Wife (July) Wife + (Dec) Military Son	M C		
1948						Jan	Cleaning Lady	Gov't Hosp.	.65 hr .80 hr	40				
1949						Dec	Sales Clerk	10¢ Store	.95 hr	40				
1950											(Dec) Wife + son	M		
1951	June divorce					June					(Jan) alone	M	L	G.I. Bill (Sept)
1952														
1953														
1954											(July) Roommate		L	
1955														(June)
1956	July married	23		yes	A.B.	July	Journalist	Daily News Paper	$5,000 yr	40	(July) Wife	C		
1957				no										
1958			F-May	-		Feb			$5,250 yr	40	Wife + daughter (May)			
1959			F-Sept	-							" + 2 daughters (Sept)			
1960				yes								S		
1961														
1962														
1963														
1964						Sept	Journalist	Daily News Paper	$7,500 yr	35	Wife, 2 daughters, Mother-in-law			
1965												C (illness)		
1966												S		
1967														
1968														
Current	married								$9,500					

Figure 1 cont.

at the beginning of a calendar year, while our procedure allows changes to take place at the beginning of any month. In principle, if a respondent held twelve one-month jobs in a given year, all twelve would be recorded. For example, consider the cases of schoolteachers who have summer jobs, individuals who take on a second full-time job for part of a year, and seasonal workers who are employed by different employers at different times of the year. In addition, knowing the month in which an event took place allows us to determine the temporal ordering of events in situations where more than one significant change took place in different "life areas" in a given year.

To insure accuracy in the questionnaires, interviewers were provided with self-adhesive "age strips" which they attached to the questionnaire after determining the calendar year in which the respondent was 14. The respondent then had the flexibility of recalling events in terms of either calendar dates or his age, and the interviewer had no difficulty entering it in the appropriate row.

Field Work

As previously noted, the target population in our study is the total population of males 30-39 years of age residing in households in the United States. Individuals in the sample were selected by standard multi-stage area probability methods. The National Opinion Research Center (NORC), University of Chicago, assigned both black and white interviewers to this survey, and whenever possible the race of the respondent and the race of the interviewer were matched. A two-stage training program was used to provide specific instruc-

TABLE 1

Summary of information collected in life history survey

A. Time-Dependent Data, Age 14 to Present

Variable Name	Details	Comments and Definitions
1. Full-time education	Name and/or type of school Degree/highest grade completed	If respondent received "public support" for education (e.g., G.I. Bill of Rights), that information is recorded under No. 11.
2. Full-time employment/ unemployment	For each period of employment, the following is recorded: (a) occupation (b) industry (c) starting and ending wages (d) wages in kind (if applicable) (e) hours/week (f) reason for termination: own decision not own decision (g) at termination: had new job knew of job neither (h) got job through: friends public agency private agency advertisement other (i) formal on-the-job-training (j) union membership	Full-time employment is defined as employment which requires at least 35 hrs/week.

TABLE 1 (continued)

3. Part-time employment	Same as (a)-(e) of No. 2 above.	Part-time employment is defined as employment which requires less than 35 hrs/wk.
4. Part-time education	Name and type/purpose of school; Diploma/certificate or course	Formal schooling in which respondent engaged on a basis that was less than the standard "full-time student."
5. Military service	Drafted or enlisted Location of <u>longest</u> stay while on active duty Place of discharge Ranks at entry and discharge Military education	Military education is education which the respondent received in the military which is equivalent to civilian education. Excludes training received for military occupations.
6. Other full-time activities	Activities not included in other "full-time" variables	E.g. major illnesses, travel, prison, etc.
7. Family history	Marital status Age of wife at marriage Births, deaths, and adoptions of children	Common Law marriages are included as "marriage."
8. Wife's education	Degree or highest grade completed at time of marriage Education subsequent to marriage	For wife's education subsequent to marriage, only the completion date recorded.
9. Wife's employment	Same as items (a)-(e) of No. 2 above	Wife's employment recorded only during marriage to respondent.

TABLE 1 (continued)

10. Household composition	Relationships to respondent of all individuals with whom he was living	
11. Income adequacy	R's perception of whether or not the household of which he was a member was: able to save live comfortably without saving just manage need outside help	If R reports that the household needed outside support, the source (if public) was recorded.
12. Home details	Type of housing No. of rooms Rent/own/share Neighborhood	For each housing unit, we inquired as to whether the immediate area was all white; mostly white; half white- half non-white; mostly non-white; or all non-white.
13. Migration history	City/town County State	If foreign, name of country only. For each rural location, we record whether location is farm or non-farm.
B. Cross-sectional data		
14. Place of birth		
15. Siblings		
16. Father's education Mother's education		If no father and/or no mother at age 14, education of father substitute recorded and the relationship to respondent of the substitute is specified.

TABLE 1 (contintued)

17. Father's occupation and industry Mother's occupation and industry		Same instructions for "substitute" apply here as in the case of parental education
18. Family happiness at age 14		
19. Nationality	"What is the main nationality in your mother's background?" "What is the main nationality in your father's background?"	Asked of whites only.
20. Religion	a) Religion of family of origin b) Present religion	If a) and b) are different--year of change is recorded
21-22. Voting behavior	a) Last presidential election b) Year of first vote	
23. Open-ended item dealing with major changes in Respondent's life		After respondent completes reply, direct probes deal with Family life Job/career Education Military service Major moves if respondent has not mentioned these life areas previously.
24. Short-form verbal achievement test		Adopted from Miner
25. Social security number		Requests number and permission for access to Social Security file.

tions for this survey. First, supervisors from each Primary Sampling Unit (PSU) were trained either via long-distance conference calls or in person by the NORC staff. Supervisors in turn trained interviewers in their PSU's, and, in addition Project staff members conducted a number of training sessions in the field.

Each interviewer was provided with questionnaire specifications prepared under the direction of the Project staff and modified by NORC. These specifications contained precise explanations and definitions for all the information to be collected and suggested appropriate probes to be used in the field. In general, interviewers were encouraged to collect the information pertaining to a specific life area in temporal sequence (either forward in time as in the case of education, or backwards as in the case of marital status by starting with present status). However, the format does allow for alternative methods of obtaining the information. For example, a respondent may report his occupation at a given point in time and relate other information to it -- where he was living, with whom he was living, etc.

Recording information across rows, starting with a column in the center of the questionnaire (e.g., family history) or using other alternatives in interviewing was not discouraged. Reports from interviewers indicate that variation in the interview situation did, in fact, exist. For example, our qualitative impression is that higher status respondents had no difficulty in reconstructing a consecutive job history abstracted from family history or residence. On the other hand, many lower status respondents continuously

linked job information to significant family events, major geographical moves, etc.

Our instructions also encouraged the participation of the wife whenever possible. We thought that her recall could be an asset in getting facts and sequence of events in proper order. In addition, the wife would have more complete information about the parts of the interview which concern her own history and the family history since the time of her marriage. Our experience also indicates that the rapport between interviewer and respondent was better when the wife was not excluded.

Each interviewer was required to complete one "live" interview prior to beginning her assignment. This interview was carefully reviewed by supervisors and, where necessary, additional training was provided. These "practice" interviews were also useful to the project staff both as a "zero hour" check on the specifications and instructions before the study actually began, and in the development of coding procedures and the testing of computer programs.[8]

In addition to the training, a "field edit" was instituted to insure quality. This consisted of a complete review of the first few interviews returned by each interviewer and a partial review of the remaining questionnaires from each interviewer's assignment. In the case that a field edit showed incomplete or inconsistent responses, the questionnaire was returned to the interviewer for further clarification by the respondent.

---

[8]Although we had conducted a number of local pre-tests, the national practice interviews alerted us to regional problems of wage rate units, definitions, etc.

Phone validation was conducted on 5% of the cases. In some instances, minor items of information which had been detected as missing during the field edit were solicited from the respondent during the validation call.

EVALUATION

As Balan et al. note, there are problems in the evaluation of life histories connected with the reliability of response and recall and the tendency of respondents to present "congruent" responses to agree more closely with the structure of the questionnaire than circumstances warrant. Hence, in both studies, a format was used to cross-reference and relate events to each other, thus minimizing problems connected with respondent error and recall. However, checks external to the interview itself can most satisfactorily resolve some of the doubts about reliability.

The external checks to be performed on our data consist of two types. First, we can compare the distributions of variables in the life history sample with documents issued by the federal government and academic publications. An example of the latter type is Blau and Duncan's (1967) analysis of their survey "Occupational Changes in a Generation." Presumably, our aggregate distributions for this cohort should not only be comparable to some of their 1962 data, but some of the retrospective data collected in that survey should also be comparable to our retrospective data (e.g., father's occupation).[9]

---

[9]For a discussion of Census checks on retrospective data, see Appendix E of Blau and Duncan (1967).

The second type of external check is a comparison of a small segment of our data (earnings) with data collected by the Social Security Administration. Each respondent was asked to give signed permission for access to his Social Security File, and by a special arrangement with the Social Security Administration, we will be able to obtain earnings information for consenting respondents for all covered employment.[10]

We recognize that there are inherent difficulties with this procedure. The earnings records of the Social Security Administration are derived from report forms submitted by employers and self-employed persons. Employers report quarterly (annually for agricultural workers) during the month following the calendar quarter (or year). Wages, however, are reported by each employer up to the maximum for the given year.[11] Furthermore, not only is not all employment represented in the earnings records, but program changes from time to time have expanded coverage.[12] Therefore, discrepancies between the quarterly (or yearly) earnings which we calculate from our interviews and those obtained from the SSA may be the result of many factors other than incorrect recall on the part of respondents.

---

[10] Permission was obtained from about 82% of the respondents.

[11] The SSA maximum has changed as follows: 1937-50: $3,000; 1951-54: $3,600; 1955-58: $4,200; 1959-65: $4,800; and 1966 and later $6,600. The SSA has developed procedures with which to estimate total earnings for an individual on the basis of quarterly reports.

[12] At present, about 90 percent of persons in paid employemnt are covered by the Social Security Administration.

PROCESSING THE DATA

While the general approach followed both in the questionnaire format and in the interview situation is similar to that of the Monterrey study, we departed substantially from the previous work in the procedures for transferring life histories from the interviews to computer tapes.

Editing and Coding

Although the cases which arrived at the University had been partially edited in the field, each case was subject to careful editing by a staff member prior to coding. Aside from checking each case for completeness, information contained in the marginal notes was incorporated into the data, foreign currency converted to dollar equivalents, etc. In order to simplify coding, a number of variables were coded on the questionnaire itself after editing: place names, union names, and occupations and industries.[13]

In transferring information from the questionnaire to coding forms the Monterrey study coded only changes in the content of a variable. For this coding, they used a large matrix form in which columns were assigned to variables and rows to each of the years covered by their survey. (Each change in the content of a variable was coded in a six-digit block. The first

---

[13] Military bases were assigned the place name of the closest community. In coding place names, we devised a procedure which will allow us to use a table look-up program with information available on the City and County Data Book tapes. To insure accuracy in the coding of occupations and industries and avoid the cost of training coders for this task, we were fortunate in having the services, on a consultant basis, of an individual with twenty years of experience with the Bureau of the Census.

two digits identified age; the third and fourth identified the variable; and the last two digits refered to the specific content of the variable.)

Our approach to coding was fundamentally different, a difference with implications not only for data processing procedures but also for the analytic utility of the data as will be noted below. Instead of coding changes, we coded the different states in which an individual was in at various times in his life. Thus, we are coding states which have time as an integral part of them, but which are not defined with respect to a fixed time matrix. This approach meant that we had to devise a coding procedure which would allow flexibility of coding as many states as the respondent reported in the interview situation. Only in focussing upon states could we preserve the richness in our data which results from recording as many statuses as the respondent reports and knowing the month in which status changes occur.[14]

In order to accomplish our aim, we used seven distinct 80-column coding forms which place no restrictions on the number of changes in a specific "life area" for a given respondent. These coding sheets were partially precoded with variable numbers and formatted to facilitate coding, keypunching and subsequent machine processing. Static variables such as birthplace, parental education, social security number, etc. were coded on the first codesheet (three cards) in a fixed format. Subsequent codesheets (2-7) were used for recording the non-static variables -- such as education, occu-

---

[14]Our procedure is applicable no matter what the time-unit of analysis happens to be. Thus, we could (if data were available) record the day in which a respondent entered and left a specific state.

pation, residence -- in which the number of recorded states is subject to individual variation. Thus, for a variable such as occupation, if the respondent held ten jobs, ten "states" (cards) would be recorded. Each card gives the starting and ending dates of the job, the starting and ending wages, the occupation and industry, plus other data pertaining to this period of employment. Since the number of entries varies for each respondent, continuation or ending flags (M's or V's) were coded in the last entry in this variable (see Figure 2).

To economize on the number of codesheets, several of the codesheets are used for recording more than one variable unit. In this case, each subsection was treated separately by the coders. Thus, for example (Figure 3), coders complete all entries for marital status (Variable 17) vertically before proceeding to the next variable (27). As elsewhere, only changed states in a variable are coded and the continuation and ending flags for each variable are coded on these multi-variable codesheets. Inspection of Figure 3 shows that the respondent occupied three marital states (Variable 17) and occupied thirteen states of residence (Variable 10).

Aside from the fundamental difference discussed above, several procedural advantages are derived from using distinct codesheets as opposed to a single matrix type coding form. Coders found it very easy to accurately transcribe the data from the questionnaire since the codesheets closely resembled the questionnaire in format. We minimized errors which can be made by having to leave blanks in the matrix form by recording consecutive columns and rows. Our procedure allowed each coder to be responsible for

CASE NUMBER 5999

Coded by WR Checked by 郭

The Johns Hopkins University

SOCIAL ACCOUNTS SURVEY

Codesheet 3

OCC/IND $99,999.00 $99,999.00

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
0	1	5	9	9	9	0	2	1	4	6	1	1	9	9	3	9	9	3															4	6	1	2														

CASE NUMBER 5999

Coded by WR Checked by JP

The Johns Hopkins University
SOCIAL ACCOUNTS SURVEY

Codesheet #5

Case #	Variable #	Continuation Marks	Variable #	Continuation Marks	Variable #	Continuation Marks	Variable #
01	5999	1	7461121	6M	2714707101	1M	3714611 0M 1043990203095 2
02	5999	1	751064	FM	2715106701	1M	3725607 1M 1045060203 52
03	5999	1	75607223	V	2715805202	2M	3725799 0M 1046110203 5207
04	5999				2715909203	2V	3726099 1V 1047060 7
05	5999						1047070771
06	5999						104712 40
07	5999						1050120771
08	5999						105101
09	5999						10540730
10	5999						10560707
11	5999						1058050781
12	5999						1059090782
13	5999						1064990782 11

Figure 3

Multi-variable Coding Form
(Family History)

coding specific variables rather than entire cases, thereby developing proficiency with that particular variable. Moreover, keypunching and subsequent cleaning and checking of the data on the computer are thereby simplified.

Machine Processing

Before any analysis of the data could be carried out, a set of special computer programs was written to clean the data and check for inconsistencies, to merge the data for each individual and finally to retrieve the information in a format usable as input to existing computer programs.

A basic way in which we differ from the Monterrey study is in the method of storing and retrieving data from magnetic tape. As discussed previously, we recorded only states, whereas the Monterrey study recorded changes; i.e., we incorporate time as a part of the specification of a state instead of recording changes as a function of time. This use of time as an interior portion of the data rather than as an exterior framework within which recordings are made is a fundamental distinction between the two approaches. Thus, our unit of time could easily become days, hours or years, or any measure of time. From their recorded changes, the Monterrey group then reconstructed (by means of a computer program) the entire time-variable matrix for every variable for every year. Essentially, in transferring data from cards to tape, they repeated the assigned code for each year until a change in the variable was encountered, then repeated the new assigned code, etc. Thus, the amount of data stored on the tape was the same for all individuals (of a given age) regardless of the number of changes recorded for him.

In actuality, we are interested only in those elements of this time-variable matrix where a new assigned code occurs (i.e. a change in status). Thus, the number of elements of interest in this matrix is small.[15] Most of the entries in the Monterrey data are repeated codes which serve as fillers to change variable format data to fixed format data. This computer constructed matrix does have the virtue of allowing direct retrieval of data from it by specifying only a row and column of the matrix. However, a sparse matrix, such as this one can most efficiently be stored as a variable length list, given of course that there is some way to retrieve data from the list.[16]

By storing the data for each individual in a variable length tape record and then providing an "index" to the record, we eliminated the need for contructing a fixed format record. This index contains, for every state of every variable, the time span of the state and the position on the master data tape of this state.[17] Figure 4 is a portion of the index dealing with a number of variables. Using this index, retrieval of a state of a variable for any date or age becomes quite easy.

---

[15]Consider, for example, a respondent aged fifty who completed grade twelve at age eighteen. In the Monterrey study, the education variables would have the same code repeated each year from ages 18-50. Thus, the proportion of data of interest from ages 18-50 is 1/32 of entire recording. As is apparent, there would be no loss of information by stopping the recording of education at age eighteen, since from 18-50 the variable does not change.

[16]Balan et al note, op cit. p. 111, that for their thirty variables (45 columns of codes) for a respondent aged 60, the resulting matrix has 2700 positions. Using their approach, had they used the month as the unit of time, it would result in a matrix of 32,400 positions. The minimum length list for our data is 683 positions and the maximum is 7,887 positions.

[17]Each date in the sample was transformed from the original month-year code to a continuous code where January, 1929, the earliest birthdate possible for anyone in the sample became 001.

Identification Code	Birthdate in Months	Variable # (Full-time Occupation)	Number of Entries this Variable	Beginning Date: Entry 1, Variable 02	Ending Date: Entry 1, Variable 02	Position on Master Tape: Entry 1, Variable 02	Beginning Date: Entry 2, Variable 02	Ending Date: Entry 2, Variable 02	Position on Master Tape: Entry 2, Variable 02	Beginning Date: Entry 5, Variable 02	Ending Date: Entry 5, Variable 02	Position on Master Tape: Entry 5, Variable 02	Variable # (Part-time Occupation)	Number of Entries	Beginning Date: Entry 1, Variable 03	Ending Date: Entry 1, Variable 03	Position on Master Tape: Entry 1 Variable 03
5999	0001	02	05	0215	0216	712	0216	0220	792	0435	0481	U32	03	02	297	313	V12

Figure 4

Portion of Index

Case # 5999

A generalized retrieval program was written by Karweit and Sørensen to retrieve subsets of data from a master tape onto a work tape. This output tape is then usable as input to existing, standard computer programs. One type of request this program allows is the retrieval of a variable at a given point in time. Several definitions of time associated with a variable were taken into account in the design of the program. First, "time" could mean the respondent's age, so that it is possible to ask, for example, what occupation was held by all men at age 22. Second, "time" could mean an actual calendar date. For instance, to compare the occupational distribution of this sample with the 1960 census, it is possible to retrieve the occupation in 1960 for all men in the sample. The duration of a state may also be requested, e.g., duration of job held at age 22, or duration of first job.

The program also retrieves information at a time point determined by a specific event. An example of contingent retrieval is the occupation held by the respondent during the month in which he was married or a specific time period before or after his marriage. In addition, we can have access to information on either the number of times a respondent has been in a given state up to a specified point in time or the amount of time spent in various states. An illustration of cumulative retrieval is the number of months the respondent has been unemployed to his thirtieth birthday. Also, retrieval of transitions between states is possible with the program where, for example, the probabilities of movement between specific occupations for one year periods is obtained.

Finally, it is possible to retrieve information by using a state as the unit of analysis, rather than use an individual. For example, one can retrieve a record for each full-time job occupied for the sample together with characteristics of individuals occupying these jobs.

It should be reiterated that after the data is retrieved by using this general program, the work tape can be used as input to any existing computer program. To illustrate the flexibility and power of this general retrieval program, a number of preliminary tables are presented on the following pages.

Summary

The survey described in this paper and the procedures we have developed for handling the data was designed to meet our own research needs. However, we are convinced that the procedures described above can have broad applications in social science research. With minimal modification of approach, we feel that it can be used for the collection and processing of data from larger samples and different population sub-groups. Subsequent research reports from our project will demonstrate the various types of analysis that are currently being carried out with this data.

Table 2. Retrieval by AGE: Percentage Distribution of Respondent's Occupation at Selected Ages[1]

ACTIVITY	RESPONDENT'S AGE[2]										
	16	18	20	21	22	23	24	25	26	28	30
OCCUPATIONAL GROUP:[3]											
Professionl, technical	0.1	0.6	1.9	2.0	4.2	6.5	7.9	9.9	11.6	14.7	16.6
MOP, exc. farm	0.4	0.6	2.0	1.7	2.7	3.0	3.5	4.7	6.2	8.5	9.2
Clerical and sales	1.5	5.5	8.0	8.3	7.9	9.3	9.9	12.7	12.5	11.5	11.5
Craftsmen	0.8	4.0	7.5	7.6	8.1	9.9	11.1	13.6	15.1	16.5	18.4
Operatives	5.0	12.0	16.9	16.7	17.8	20.9	23.2	21.2	22.2	21.8	19.9
Service workers	1.8	2.4	2.1	1.9	1.8	2.1	2.7	3.3	4.0	3.6	3.3
Laborers, exc. farm	6.0	8.2	11.0	9.0	9.5	9.5	9.7	9.2	7.8	7.6	7.3
Farmers and farm managers	1.2	1.4	2.1	2.3	2.1	3.0	4.0	4.5	4.3	4.6	4.3
Farm laborers and foremen	8.4	9.7	5.4	4.4	4.2	3.0	3.3	2.9	2.5	2.7	2.6
Armed Forces	0.8	9.4	22.2	28.2	25.6	18.3	10.8	5.0	2.4	0.9	0.2
Other[4]	74.0	46.3	21.0	17.9	16.0	14.4	14.0	12.9	11.3	7.5	6.6

1. Table based on Sample A, males 30-39 years of age residing in households in the United States, N = 953. See Appendix A for details.
2. Time point here is the birth month of the year in which the respondent turned the given age.
3. Includes only respondents who report working at an occupation "full-time." Full-time is defined as working 35 hours or more.
4. Includes respondents who are unemployed, working only part-time, and not in the labor force.

Table 3. Retrieval by CALENDAR DATES: Percentage Distribution of Respondent's Occupation at Selected Calendar Dates[1]

ACTIVITY	DATES							
	AUG 1948	JAN 1952	AUG 1957	APR 1960	MAR 1962	NOV 1964	JUL 1966	DEC 1968
OCCUPATIONAL GROUP: [2]								
Professional, technical	.2	.9	5.9	11.4	14.9	16.1	17.1	17.6
MOP, exc. farm	.3	.8	3.9	6.4	8.4	10.0	10.9	13.1
Clerical and sales	3.0	4.6	9.1	12.0	11.9	12.2	11.2	11.3
Craftsmen	2.1	4.2	11.5	14.4	15.9	18.5	19.1	19.5
Operatives	6.4	9.9	20.4	23.0	20.5	19.9	20.1	18.8
Service workers	2.2	1.6	2.5	3.1	2.8	3.8	3.7	3.5
Laborers, exc. farm	5.5	5.9	9.4	8.2	6.9	6.3	5.8	4.9
Farmers and farm managers	.9	1.3	3.1	3.8	4.2	4.5	4.3	4.2
Farm laborers and foremen	7.3	5.9	3.7	2.9	2.4	1.9	2.1	1.7
Armed Forces	3.8	16.9	13.1	6.1	5.0	2.6	1.9	1.4
Other [3]	68.2	48.0	17.3	8.7	7.0	4.3	3.8	3.9

1. Table based on Sample A, males 30-39 years of age residing in households in the United States, N = 953. See Appendix A for details.
2. Includes only respondents who report working at an occupation "full-time." Full-time is defined as working 35 hours or more.
3. Includes respondents who are unemployed, working only part-time, and not in labor force.

Table 4. Retrieval by DURATION: Mean Duration of Full-time Job at Selected Ages

(Includes only Respondents who were in the Civilian Labor Force at the Selected Age.)

AGE	DURATION (in Months)[1]		Respondents in C.L.F.	% of Total Sample in C.L.F.[2]
	Mean	Std. Dev.		
15	27.12	33.94	171	17.9
20	41.56	52.76	791	83.0
25	69.72	74.70	871	91.4
30	79.50	73.98	920	96.5
35	87.12	91.36	497	96.5

1. If respondent is occupying a given full-time job at interview date, duration is measured to that time-point.

2. For ages 15 - 30, N = 953.
   For age 35, N = 515.

Table 5. CONTINGENT Retrieval: Occupational Distribution of National Sample of Men 30-39 Years Old in 1968 at Five Years after First Marriage[1]

ACTIVITY	NUMBER	%
OCCUPATION[2]		
Professional, technical	122	15.0
MOP, exc. farm	60	7.4
Clerical and sales	106	13.1
Craftsmen	139	17.1
Operatives	189	23.3
Service workers	27	3.3
Laborers, exc. farm	69	8.5
Farmers and farm managers	34	4.2
Farm laborers and foremen	17	2.1
Armed Forces	36	4.4
UNEMPLOYED	12	1.5

1. Excludes men who have never been married, N = 811.

2. Includes only respondents who report working at an occupation "full-time." Full-time is defined as working 35 hours or more. This Table is based on Sample A, males 30-39 years of age residing in households in the United States. See Appendix A for details.

Bibliography

Balan, Jorge et al. "A Computerized Approach to the Processing and Analysis of Life Histories Obtained in Sample Surveys," Behavioral Science vol. 14, no. 2 (March, 1969), 105-120.

Blau, Peter M. and Duncan, Otis Dudley. The American Occupational Structure. New York: John Wiley & Sons, 1967.

Coleman, James S. Race and Social Change. New York: Random House-Alfred A. Knopf, Inc., forthcoming.

Johnstone, John W.C. and Rivera, Ramon J. Volunteers for Learning. Chicago: Aldine Publishing Co., 1965.

Appendix A

Sample Design and Selection

The target universe in the life history survey is the total population of males 30-39 years of age residing in households in the United States (henceforth designated as eligible males). Individuals in the sample were selected by standard multi-stage area probability methods.

The analytical design required approximately equal numbers of black and white eligible males. Hence the selection consisted of two parts: (1) Sample A, designed to yield the required number of white eligibles plus a number of eligible Negroes proportional to their representation in the population as a whole; and (2) Sample B, a supplementary selection of Negro households only, designed to supply the additional eligible blacks required to satisfy the analytical design.

Sample A was based upon the NORC 1962 Master Sample, a self-weighting selection of U.S. households in which sampling units at successive stages before the final stage are selected with probabilities proportional to their estimated 1965 populations. The primary sampling units (PSU's) consist of 72 standard metropolitan statistical areas (SMSA's) and rural counties selected from strata based upon classification by geographic region, size of largest town, median family income, economic characteristics and in the South, race. Additonal details of selection procedures within PSU's, involving the selection of localities, tracts, enumeration districts, city blocks, and in some cases directory

sampling, are provided in the appendices of several NORC publications (e.g., Johnstone and Rivera (1965)).

Sample B was designed to obtain the additional black eligibles without the massive screening of households that would be necessary in a national sample with selection proportional to overall population size. In order to minimize the cost of setting up new field staffs, the same PSU's were employed in Sample B as in Sample A, with the addition of two PSU's in the South aimed at reducing the excessive clustering of eligible Negroes in rural areas. After the primary stage, however, the selection of sampling units was made with probabilities proportional to the estimated Negro population instead of total population. At the final stage, segments of households were selected with equal probabilities and screened for black eligibles who were interviewed upon identification. The end result of this procedure was a self-weighting sample of black eligibles only, with the variation in cluster size kept within efficient bounds by the use of probabilities proportional to Negro population size.

The overall probability of selection in Sample B was 6.639 times that in Sample A. Furthermore, the same probability of selection applies to the Negroes in Sample A since they also had an opportunity of falling into Sample B if not selected in Sample A. Hence, in combining cases from both samples for the purpose of estimating characteristics of the target population, the white observations from Sample A receive a relative weight of 6.639.